November 2023

Automatic Dish Name Extraction from User-generated Content Using LLM

Bo Lin

Johann Hibschman

Kathleen Oshima

Automatic Dish Name Extraction from User-generated Content Using LLM

ABSTRACT

Extraction of dish names from user-provided content such as food photographs and captions, restaurant reviews, and other free-form text is a challenging task. Rule-based approaches are difficult to maintain and improve. Pattern matching against a predefined dictionary often suffers from low recall. Conventional machine learning models require large amounts of labeled data to perform named entity recognition (e.g., to recognize dish names) which is often costly and does not scale well across multiple languages and countries. This disclosure describes the use of a multimodal large language model to automatically extract dish names from user-generated content such as food photographs and associated free-form text such as tags, captions, etc. Dish name extraction from the user-provided tags can be formulated as an open vocabulary dish name entity recognition and discovery task, which fits naturally with the framework of pre-trained LLMs, and leverages the model capability in handling multilingual, multicultural text understanding.

KEYWORDS

- Named entity recognition
- Dish name
- Large Language Model (LLM)
- Multimodal LLM
- Multitask Unified Model (MUM)
- User-generated content
- Restaurant menu
- Food recommendation

BACKGROUND

Many users upload photographs of various dishes after visiting a restaurant and attach tags or captions for the photographs. Such photographs can be viewed in digital maps, restaurant search/database applications, etc. Tags and other content associated with such photos can be used for food recommendations to other users that view the restaurant (e.g., “dish A is popular here”) or to help users in the search for a particular dish or combination of restaurant and a dish name.

Since user-provided tags are noisy free form text across a variety of languages and in different countries, the dish name understanding task is challenging. In one approach, a dish dictionary can be used in a pattern matching technique to recognize dish name mentions from the user-provided free-form text. This approach can also benefit from the use of the menu information from available restaurant menus. However, this can still suffer from low recall. Given that new dish names are invented constantly across the restaurants in the world, the size of the dictionary required to capture all the possible dish names across multiple languages and for multiple countries can get prohibitively large. Another technique is to use an open-vocabulary approach where dish name mentions are identified from the text by leveraging a machine learning model that is trained on a large amount of labeled text. This approach suffers from the high cost associated with the labeling task and the lack of high quality label data for resource-scarce languages or countries. Besides machine learning based approaches, rule-based approaches for dish name extraction are even more difficult to maintain and improve. None of the approaches is suitable for large scale dish name extraction across multiple languages and countries.

DESCRIPTION

This disclosure describes the use of a multimodal large language model to automatically extract dish names from user-generated content such as food photographs and associated free-form text such as tags, captions, etc. Dish name extraction from the user-provided tags is formulated as an open vocabulary dish name entity recognition and discovery task. The task fits naturally with the framework of pre-trained LLMs, and leverages the model capability in handling multilingual, multicultural text understanding.

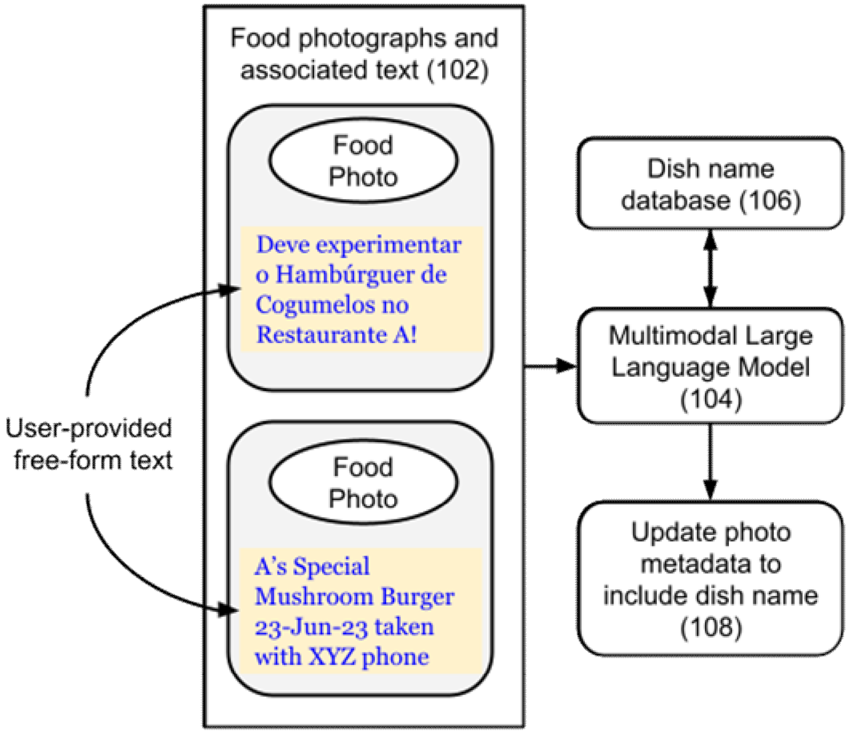
Food photographs and associated text (102)
Food Photo
Deve experimentar o Hambúrguer de Cogumelos no Restaurante A!
Food Photo
A's Special Mushroom Burger 23-Jun-23 taken with XYZ phone
User-provided free-form text
Dish name database (106)
Multimodal Large Language Model (104)
Update photo metadata to include dish name (108)

Fig. 1: Dish name identification using large language model

Fig. 1 illustrates an example of identifying dish names from food photographs and associated free-form text content, e.g., obtained from restaurant reviews or other sources, using a large language model. A large dataset of food photographs and associated text (102) is provided to a multimodal large language model (LLM) (104). A dish name database (106) is a dictionary that can be used to extract dish names from the text.

The task for the LLM is to extract dish names from the input. For example, given a photo caption "Try the mushroom burger," the task for the model is to extract "mushroom burger." This is different from extracting common dish entities from a piece of text. Extraction of dish names can enable identification of the dishes served and observed at the particular restaurant, including items that are unique to the place (e.g., “Bob’s pizza”) as well as items that have common names but are presented differently across places such as “House special noodle soup.” The extracted dish name can be associated with the respective photograph, e.g., by adding the dish name to the photograph as a label or metadata.

New dish names detected by the LLM can be added to the database. The multimodal LLM can identify dish names across multiple languages and can also translate dish names to a particular language as necessary. Dish name detection can be performed periodically as new photographs and associated text become available, e.g., as restaurants update their menus and/or as users provide new content.

For efficiency, a dictionary-based lookup may be performed initially based on candidate dish names extracted from the dataset. Any previously recognized names (present in the dictionary) are automatically identified. The multimodal LLM can then be used only to recognize dish names that are new (not present in the dictionary). The dictionary can be updated automatically. In some implementations, the dish names generated by the LLM may be constrained to be substrings of the input text to avoid model hallucination and/or unwanted translation. Further, the LLM can also be used to automatically translate dish names.

The extracted dish names can be used for various purposes, e.g., to show popular or new dishes at a particular place, to identify trending dishes, to provide recommendations, in search applications, etc.

CONCLUSION

This disclosure describes the use of a multimodal large language model to automatically extract dish names from user-generated content such as food photographs and associated free-form text such as tags, captions, etc. Dish name extraction from the user-provided tags is formulated as an open vocabulary dish name entity recognition and discovery task. The task fits naturally with the framework of pre-trained LLMs, and leverages the model capability in handling multilingual, multicultural text understanding.

REFERENCES

1. "Find the perfect dish, no matter your craving," available online at https://blog.google/products/search/food-restaurant-search/, accessed Oct 24, 2023.
2. "MUM: A new AI milestone for understanding information" available online at https://blog.google/products/search/introducing-mum/, accessed Oct 24, 2023.